DOCUMENT RESUME

ED 073 144 TM 002 390

AUTHOR	Beers, Joan S.
TITLE	Educational Evaluation.
INSTITUTION	Educational Testing Service, Princeton, N.J.; ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	ERIC-TM-16
PUB DATE	Feb 73
NOTE	8p.; 1972 AERA Conference Summaries
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Conference Reports; *Educational Research; *Evaluation Methods; *Mathematical Models; *Program Evaluation; Speeches

ABSTRACT

The 50 papers on educational evaluation presented at the 1972 AERA Conference are reviewed. The papers were classified into four categories: Theory and Methodology, Empirical Models, Empirical Studies, and Nationwide Evaluation and State Assessment. A list of the papers reviewed, their authors, and, when applicable, the ED numbers concludes the summary. (DB)

ED 073144

ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, & EVALUATION
EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY 08540

Conducted by Educational Testing Service in Association with Rutgers University Graduate School of Education

TM Report 16

February 1973

1972 AERA Conference Summaries

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EDUCATIONAL EVALUATION

Joan S. Beers
Pennsylvania State Department of Education

TM 002 990

This publication was prepared pursuant to a contract with the National Institute of Education, U. S. Department of Health, Education and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, represent official National Institute of Education position or policy.

PREVIOUS TITLES IN THIS SERIES

1. Developing Criterion-Referenced Tests
   ED 041 052

2. Test Bias: A Bibliography
   ED 051 312

3. Ability Grouping: Status, Impact, and Alternatives
   ED 052 260

4. Developing Performance Tests for Classroom Evaluation
   ED 052 259

5. Tests of Basic Learning for Adults: An Annotated Bibliography
   ED 058 274

6. State Educational Assessment Programs: An Overview
   ED 058 309

7. Criterion Referenced Measurement: A Bibliography
   ED 060 041

8. Measures Pertaining to Health Education: I. Smoking
   ED 060 042

9. Measures Pertaining to Health Education: II. Drugs. An Annotated Bibliography.
   TM 002 078 (ED Number not yet available)

10. Measures Pertaining to Health Education: III. Alcohol. An Annotated Bibliography
    TM 002 079 (ED Number not yet available)

11. 1971 AERA Conference Summary
    I. Evaluation: State of the Art
    ED 060 043

12. 1971 AERA Conference Summary
    II. Criterion Referenced Measurement
    ED 060 134

13. 1971 AERA Conference Summary
    III. Educational Statistics
    ED 060 133

14. 1971 AERA Conference Summary
    IV. Test Development, Interpretation, and Use
    ED 060 135

15. 1971 AERA Conference Summary
    V. Innovations in Measurement
    ED 060 044

INTRODUCTION

About 700 of the 1,000 papers presented at the 1972 AERA Annual Meeting in Chicago, Illinois were collected by the ERIC Clearinghouse on Tests, Measurement, and Evaluation (ERIC/TM). ERIC/TM indexed and abstracted for announcement in *Research in Education (RIE)* 200 papers which fell within our area of interest–testing, measurement, and evaluation. The remaining papers were distributed to the other Clearinghouses in the ERIC system for processing.

Because of an interest in thematic summaries of AERA papers on the part of a large segment of ERIC/TM users, we decided to invite a group of authors to assist us in producing such a series based on the materials processed for RIE. Four topics were chosen for the series: Criterion Referenced Measurement, Evaluation, Statistics, and Test Construction.

Most papers referred to in this summary may be obtained in either hard copy or microfiche form from:

ERIC Document Reproduction Service (EDRS)
P.O. Drawer 0
Bethesda, Maryland 20014

Prices and ordering information for these documents may be found in any current issue of *Research in Education.*

EDUCATIONAL EVALUATION

Joan S. Beers

If the 50 papers on educational evaluation, presented at the 1972 American Educational Research Association meeting in Chicago, are indicative of the happenings in the field, clearly something occurred in educational evaluation this past year. The most obvious occurrence is that the number of papers nearly doubled from the 26 presented at the 1971 AERA meeting.

In 1971, evaluation was still going through the defining stage. The Phi Delta Kappa National Study Committee on Evaluation report, *Educational Evaluation and Decision Making*, was hot off the press and many of the 1971 presenters directed their attention to a review of the report. Some models for implementation were developing, but few papers described actual program evaluations.

This year, the largest number of presenters (18) described evaluation models, and as many presenters (14) reported on empirical studies as they did on theoretical issues. The thinking of the PDK National Study Committee still loomed large as did the thinking of the big three S's: Scriven, Stake, and Stufflebeam.

Considering the number of papers and the space allotted for this review, the most this reviewer will attempt to do is to bring some order to the many and diverse topics included under educational evaluation, cite the author(s), and comment briefly on the contents. There will be no attempt to pass judgment on the quality of the ideas presented.

The papers fall, or in some cases, were forced, into four categories: Theory and Methodology, Empirical Models, Empirical Studies and Nationwide Evaluation and State Assessment.

THEORY AND METHODOLOGY

Fourteen papers were presented in this category. Merriman designates and defines three areas which evaluators can serve in the public school system: developing evaluation and accountability systems, providing process and product evaluation of federally funded projects, and providing training and consultative services to faculty, parents and students. He cites, also, some basic problem areas of evaluation: credibility, threat, lack of personnel, and limited resources.

Womble labels public school research a "two-faced profession." Researchers have the responsibility to find possible solutions to current educational problems and, at the same time, have the responsibility to communicate their findings to people other than trained researchers. She lists recommendations for public school researchers to be "two-faced" in order to have maximum impact on the advancement of education as well as the state of the art.

Walker examines, in light of an educational setting, the five problem areas of educational evaluation identified by the PDK National Study Committee on Evaluation: definition, decision making, values and criteria, levels problems, and the research model. He makes suggestions for their recognition and avoidance and concludes with a series of hypotheses for further investigation.

Briggs reminds us that the quality of local school evaluation is still at a low level and states firmly that increased participation of external evaluators from corporations, consulting firms and universities in the evaluation efforts of school districts will only prolong a condition that needs radical changing. He proposes that new infusions of money, a broader definition of evaluation and an administrative restructuring of evaluation activities can change the system.

Ashburn, in contrast to Briggs, sees the role of the external evaluator as necessary to bridge the credibility gap between public schools and public and private funding agencies.

Woodbury and Jacobson make a posteriori recommendation for the evaluation of a performance contract. Among the recommendations are that the evaluator be involved in all planning and decisions from the very beginning, that teachers should not administer pretests,

and that criterion referenced tasks based upon locally developed goals and objectives be used to measure pupils' performance.

Ashburn, in a second paper, introduces the idea of certification for evaluators. After analyzing the pros and cons of certification, he concludes that school districts would be served best by a certification process that involved multiple levels, required periodic updating, was based on proficiency levels, and allowed school districts to participate in the training.

Hutchinson expresses the notion that an evaluation methodology which does not assure that the data will be used does not accomplish its main purpose: to provide data for decision making. He says that for evaluation to accomplish its main purpose, the goals evaluated should be the decision maker's, the variables measured should be those of concern to the decision makers, the techniques used should possess decision maker validity and data analysis should be made comprehensible to decision makers.

Jones, expanding the views of Hutchinson, discusses further the aspect of "completeness" in evaluation. In order to attain completeness, the evaluator first must elicit the decision maker's entire goal intent. Jones describes a methodology, originated by Hutchinson, for the evaluator to lead the decision maker to the completeness of a goal intent.

Forsyth analyzes Dyer's student change model of an educational system—a model in which performance indicators are computed based upon four groups of variables: input, process, and hard-to-change and easy-to-change surrounding conditions. Reliability is the focus of attention in the paper.

Fortune identifies six problems peculiar to evaluators of social action programs: (1) the presence of value conflicts and the absence of procedures through which these conflicts can be negotiated; (2) the difficulty in identifying important baseline information needs at the beginning of a developmental program; (3) unreasonable goals; (4) the necessity to measure short-term symptoms of hypothesized long-term effects; (5) the lack of knowledge about the most appropriate time for measuring treatment effects; and (6) the inability to analyze the resulting complex interactions.

The ever eloquent Stake asks this question of the evaluator of an instructional program: "Which is more important: to tell of some very special things about the program or to provide the most veridical portrayal of the program?" He opts for giving the client a substantive portrayal of the program rather than a focus on the more prominent features. According to Stake, "if the program glows, the evaluation should reflect some of it. If the program wobbles, the tremor should pass through the evaluation report."

Tatsuoka disagrees with Stake, as well as with Stufflebeam and with Guba. He argues that randomization, experimental design, and generalizability *can* be applied to the septic conditions of the classroom. He then proposes a strategy for evaluating nationwide intervention programs such as Headstart and Title I.

Rippey discusses transactional evaluation. "... Transactional evaluation looks at the effects of changed programs on the changers themselves—on the incumbents of the roles in the system undergoing the change." Changes often involve threats to the roles of incumbents in an organization and changing programs require new skills and new behaviors. The aims of transactional evaluation are to transform the conflict energy associated with change into productive activity and to clarify the roles of all persons involved in program changes.

EMPIRICAL MODELS

Eighteen papers were presented in this category. Klein devised a formula to help decision makers compare the effectiveness of differing instructional programs. The formula is based on the rationale that general program effectiveness will increase if one or more of the following variables increases: number of objectives, success on the objectives, relative importance of the objectives, number of students in the program; or if pupil time and/or program costs decrease.

Russell and Leithwood present, in great detail, a model to help decision makers base adopt-adapt-reject decision about educational innovations on precise evaluation data.

Fisher and Ward developed a design for evaluating educational programs for culturally disadvantaged children based upon the Piaget and Inhelder Taxonomy of Human Development.

Smith presents a three-dimensional model for summative evaluation of aesthetic education programs. One dimension consists of six fine art forms; the second dimension encompasses pupils' modes of behaving and experiencing; the third dimension is affective involvement.

Jacobs reports on a four-stage model for program development and evaluation at the local school level. The fundamental thrust of the model is for more educational programming to be initiated at the local school level.

Roid describes models for course evaluation in colleges and universities based upon systems designs. He observes, however, that there is little evidence that universities

reward thorough evaluation of their courses by professors. He concludes that the important task is not the presentation of new systems or models but the changing of the structure and priorities of the university towards accountability.

Doherty puts evaluation within the framework of PPBS, with educational goals forming the basis for all programs.

Fraley developed an instructional accomplishment index:

$$\text{Instructional Accomplishment} = \frac{\text{Learning}}{\text{Dollars Cost x Learner Time}}$$

to evaluate instructional modules. He details, with data, the use of the index.

Lasser reports on a model for resource support in the development of exportable instructional products. Defined and operationalized at the Southwest Regional Laboratory, the model distinguishes several stages of continuing development effort and five broad functional areas of staff assignment.

Doyle presents a model for doing transactional evaluation in program development and an evaluation design for the Ford Training and Placement Program for teachers at the University of Chicago.

Hess and Wright list five stages through which curriculum development projects typically move, identify five different audiences for the information acquired by evaluative activities, and identify five major dimensions of a comprehensive evaluation of curriculum products.

Light presents specified procedures for evaluating materials during their in-content tryout. She concludes that systematic formative evaluation is feasible even though classical experimental designs are not practical in formative evaluation. The systematic elimination of rival hypotheses is one design which appears useful in identifying inadequacies within an instructional system and in generating appropriate revisions.

Johnson reports a general conceptual model of educational research and development incorporating evaluation processes used in planning the National Program on Early Childhood Education at CEMREL, Inc. This model includes both formative and summative evaluation activities.

Bashook reveals an attempt to bridge the gap between theory and practice in teaching. The paper describes an exploratory study to devise a method to analyze and evaluate concept teaching in university science courses.

Luft, Lujan and Bemis describe the Quality Assurance Model for process evaluation developed by the Southwestern Cooperative Educational Laboratory. The model provides administrators with the opportunity to maximize desired outcomes of educational programs.

Abedor describes the development and field testing of a flow chart model for formative evaluation of self-instructional multi-media learning systems.

Miller reviews the development and application of a formative evaluation model in the design of a mathematics laboratory for young children.

Finally, Huberty presents an evaluation system for a psychoeducational treatment program for emotionally disturbed children. He emphasizes that it is important for the evaluation to be easily implemented and clinically useful.

EMPIRICAL STUDIES

Fourteen papers are in this category. Cypress and DeBloois describe the application of a formative evaluation procedure for school staffing models. Three models were evaluated as to comprehensiveness, feasibility, and viability and classified according to four organizational types. The authors conclude that the evaluation techniques have potential for evaluating the characteristics of any school organization.

Otto describes one school district's application of the System for Objective Based Evaluation–Reading, developed by the Center for the Study of Evaluation at UCLA. To date, the district established goals and performance objectives for each grade level and began development of an assessment system.

VanMondfrans, Schott, and French compared students' achievement in a variety of subject areas under block scheduling and traditional scheduling. Overall results in subject matter tests favored students in the traditional scheduling treatment. There were no significant differences between treatments for attitude and interest scores.

Kievit details an evaluation of a training program to prepare teachers to serve as workshop leaders to initiate curriculum change. The findings support the feasibility of selecting teachers for leadership training who are most likely to be responsive to efforts to diffuse innovation.

Rush, McElhinney, and Junkel evaluate the impact resulting from training and engaging public school educators as data collectors. Based on data collected through participant observations and questionnaires, the investigators concluded that public school personnel can be trained to serve effectively as data collectors for curriculum evaluation.

Barry reports the results of an inservice program on evaluation for elementary principals and curriculum coordinators. Although the results show that participants scored significantly higher than the control group in cognitive skills, only four of the 24 participants scored above 70 percent on the cognitive test. The test is included in the paper.

Melching describes the results of an effort to apply a classification system to a set of terminal objectives in reading. He concludes that reading objectives are no more difficult to classify than are objectives in other instructional content areas.

Hartlage reports the results of a comparative evaluation of three approaches to initial reading instruction–the phonic approach, a look-say approach, and a special alphabet approach. The data suggest that for beginning first graders whose readiness levels are at or above the national average, the special alphabet approach is significantly better for teaching word recognition.

Ginther conducted transactional evaluation to learn more about what teachers emphasize in their work with senior medical students in a group clinic. Naturalistic observational techniques were used to gather data.

Dembo and Wilson report on an evaluation of a performance contract in reading. Only 13 per cent of the 2,500 seventh grade pupils reached the objectives of the program. The authors make several recommendations to school districts who are involved in performance contracting.

Ellner describes the results of a summative evaluation of a program to prepare day care administrators. As a result of instruction, 75 per cent of the objectives were achieved. The paper provides a model for curriculum development and evaluation of much-needed day care training projects.

The final two papers in this section are ambitious and exhaustive studies of Title I programs and evaluation practices. Hayman, Mazure and Napier report on a survey of Title I evaluation practices in 20 urban districts throughout the country. Five general areas of concern are identified: planning and funding, design and implementation, imparting the decision process, personnel, and state and federal relationships. One resulting recommendation is that at least five per cent of project resources should go into evaluation.

Brown conducted a study of the composition and disbursement of 16 Title I projects assigned to 63 schools in Philadelphia. His design included correlational analysis, factor analysis, and content analysis. One of the findings suggests that although the needs of disadvantaged pupils traditionally have been combined under a general term "target population," significantly different subsets of pupils and schools exist within a LEA such that general–needs, pupil–service needs and achievement patterns of each subset represent a distinctly different variety and level of resource funds. He suggests that evaluation of Title I projects be reconsidered to provide data relevant to particular subsets of pupils.

NATIONWIDE EVALUATION AND STATE ASSESSMENT

The final four papers are in this category. Willard speaks to the question: "What kind of evaluation needs has the Federal government, particularly the USOE?" He proposes the nationwide survey as one approach to nationwide evaluation and describes the structure for such a survey as developed by the Joint Federal/State Task Force on Evaluation. He recommends that the survey approach be used only as a means to answer policy issues that call for a few simple questions. For policy issues that require more complex data, he suggests alternative approaches, such as observation techniques.

Bickner and Mood highlight some of the problems in translating research findings into educational policy at the national level. They discuss the divisions of responsibility for education, the multiplicity of educational objectives, the lack of faith in research findings and the consequences of rapid change.

Thorndike talks about some aspects of the results from a study of reading achievement in 15 countries. The inability to differentiate the effects of different types and qualities of schools on achievement–effects apart from pupils' family background–is not only a national problem but also an international problem.

Impara reports on educational assessment in the state of Florida. The Florida plan calls for assessment not only in the areas of basic skills but also in the areas of communication and learning skills, citizenship, occupational interests, mental and physical health, home and family relationships, esthetic and cultural appreciations and human relations.

A Final Note

If the purpose of educational evaluation is to provide information for decision making, the ultimate usefulness of evaluation is yet to be recorded. What kinds of decisions are made as a result of evaluation? How useful is evaluation to the decision making process? If educational evaluation is to continue to move forward, hopefully the answer to these questions will be one focus of the papers to be presented at the 1973 AERA meeting.

PAPERS REVIEWED

Abedor, A.J. Development and validation of a model for formative evaluation of self-instructional multi-media learning systems. 44p. (ED 064 327, MF and HC available from EDRS.)

Ashburn, A.G. The credibility gap between public schools and funding agencies. 16p. (ED 065 604, MF and HC available from EDRS.)

Ashburn, A.G. The local school district's view of evaluator certification. 10p. (ED 064 366, MF and HC available from EDRS.)

Barry, M. A study of an inservice program of evaluation in Escambia County, Florida Public Schools. 39p. (ED 062 390, MF and HC available from EDRS.)

Bashook, P.G. A methodology for analyzing and evaluating teaching strategies in university science teaching. 45p. (ED 064 375, MF and HC available from EDRS.)

Bickner, R.E., & Mood, A. Some problems associated with nationwide evaluation and the formulation of educational policy. 11p. (ED 060 568, MF and HC available from EDRS.)

Briggs, P.G. Putting the research into school district research and evaluation efforts. 16p. (ED 064 313, MF and HC available from EDRS.)

Brown, E.K. A study of ESEA, Title I impact components on urban elementary schools and their pupils. 89p. (ED 061 253, MF and HC available from EDRS.)

Cypress, B.K., & DeBloois, M.L. A paradigm for the evaluation of school staffing models: An empirical study. 48p. (ED 061 275, MF and HC available from EDRS.)

Dembo, M.H., & Wilson, D.E. An evaluation of a performance contract. 28p. (ED 064 387, MF and HC available from EDRS.)

Doherty, V.W. PBBES and school system evaluation—A progress report. 13p. (ED 061 278, MF and HC available from EDRS.)

Doyle, W.J. The role of transactional evaluation in program development. 41p. (Document not yet available from EDRS.)

Ellner, C.L. Summative evaluation of a curriculum developed to prepare day care administrators using mastery/non-mastery criteria. 35p. (ED 065 553, MF and HC available from EDRS.)

Fisher, M.D., & Ward, V.S. A design for evaluating educational programs for culturally disadvantaged children. 17p. (ED 061 281, MF and HC available from EDRS.)

Forsyth, R.A. Considerations related to the usefulness of the performance indicators in Dyer's student change model of an educational system. 21p. (ED 063 332, MF and HC available from EDRS.)

Fortune, J.C. Problems facing the evaluator of social action programs. 9p. (ED 064 380, MF and HC available from EDRS.)

Fraley, L.E., & Vargas, E.A. The measurement of instructional accomplishments. 24p. (ED 064 386, MF and HC available from EDRS.)

Ginther, J.R. Transactional evaluation in a medical school setting. 9p. (ED 062 904, MF and HC available from EDRS.)

Hartlage, L.C. Does it matter which initial reading approach is used? 2p. (ED 061 277, MF and HC available from EDRS.)

Hayman, J.L., & Others. Critical problems of evaluation in the great city schools: Diagnosis, analysis, suggested new directions. 97p. (ED 064 332, MF and HC available from EDRS.)

Hess, R.J., & Wright, W.J. Evaluation strategies as a function of product development stages. 30p. (ED 064 364, MF and HC available from EDRS.)

Huberty, C.J. An evaluation system for a psychoeducational treatment program for emotionally disturbed children. 28p. (ED 064 353, MF and HC available from EDRS.)

Hutchinson, T.E. Some overlooked implications of the purpose: To provide data for decision making. 7p. (ED 063 327, MF and HC available from EDRS.)

Impara, J.C. A system of educational assessment in the state of Florida. 8p. (ED 063 335, MF and HC available from EDRS.)

Jacobs, J.N. A model for program development and evaluation at the local school level. 27p. (ED 061 269, MF and HC available from EDRS.)

Johnson, T.J. Evaluation in the context of product and program development in laboratory and research and development centers. 121p. (ED 065 541, MF and HC available from EDRS.)

Jones, L.A. methodological gap–The completeness of a goal intent. 12p. (ED 063 326, MF and HC available from EDRS.)

Kievit, M.B. Investigation of the effectiveness of a design to initiate curriculum change. 24p. (Document not yet available from EDRS.)

Klein, S. Procedures for comparing instructional programs. 24p. (ED 061 271, MF and HC available from EDRS.)

Lasser, B.R. Resource support in the development of exportable instructional products. 9p. (Document not yet available from EDRS.)

Light, J.A. Formative evaluation procedures for the in-context development of instructional materials. 35p. (ED 065 557, MF and HC available from EDRS.)

Luft, M., & Others. The quality assurance model for process evaluation. 25p. (ED 065 556, MF and HC available from EDRS.)

Marshall, J., & Draper, A. Evaluation: Vocabulary Development Project, 1969-1970. 55p. (Document not yet available from EDRS.)

Melching, W.H. Evaluation of terminal objectives in reading. 12p. (ED 061 274, MF and HC available from EDRS.)

Merriman, H.O. The purpose and problems of public school departments of evaluation. 10p. (ED 060 541, MF and HC available from EDRS.)

Miller, G.R. The use of formative evaluation procedures in the development of a mathematics laboratory. 25p. (ED 063 341, MF and HC available from EDRS.)

Otto, R.C. Ober: A program for objective based evaluation in reading at the Newport-Mesa Unified School District. 21p. (ED 064 325, MF and HC available from EDRS.)

Rippey, R. Introduction: What is transactional evaluation? 6p. (ED 060 071, MF and HC available from EDRS.)

Roid, G.H. Systems design for course evaluation. 43p. (ED 064 398, MF and HC available from EDRS.)

Rush, D.E., & Others. Cooperative curriculum evaluation: Application of a theoretical curriculum evaluation model. 10p. (ED 060 569, MF and HC available from EDRS.)

Russell, H.H., & Leithwood, K.A. Evaluation and the diffusion of educational innovations. 48p. (ED 063 330, MF and HC available from EDRS.)

Smith, L.M. The use of observational strategies for the development of a summative evaluation model. 12p. (ED 064 367, MF and HC available from EDRS.)

Stake, R.E. An approach to the evaluation of instructional programs. 8p. (ED 064 350, MF and HC available from EDRS.)

Tatsouka, M.M. Nationwide evaluation and experimental design. 16p. (ED 064 412, MF and HC available from EDRS.)

Thorndike, R.L. Reading comprehension across national boundaries. 7p. (ED 064 351, MF and HC available from EDRS.)

VanMondfrans, A.P., & Others. Comparing block scheduling and traditional scheduling on student achievement and attitudes. 10p. (ED 064 369, MF and HC available from EDRS.)

Walker, J.P. Installing an evaluation capability in an educational setting: Barriers and "caveats." 22p. (ED 063 339, MF and HC available from EDRS.)

Willard, R.W. The nationwide survey. 10p. (ED 064 404, MF and HC available from EDRS.)

Womble, M.L. Public school research–The two-faced profession. 8p. (ED 061 272, MF and HC available from EDRS.)

Woodbury, C.A., Jr., & Jacobson, M.D. Recommendations for the evaluation of performance contracting: And a posteriori view. 4p. (ED 063 337, MF and HC available from EDRS.)